W246–W251 *Nucleic Acids Research, 2014, Vol. 42, Web Server issue* *Published online 21 April 2014*
doi: 10.1093/nar/gku291

AlignMe—a membrane protein sequence alignment web server

Marcus Stamm†, René Staritzbichler†, Kamil Khafizov and Lucy R. Forrest*

Computational Structural Biology Group, Max Planck Institute of Biophysics, Frankfurt am Main 60438, Germany

Received January 10, 2014; Revised March 17, 2014; Accepted March 30, 2014

ABSTRACT

We present a web server for pair-wise alignment of membrane protein sequences, using the program AlignMe. The server makes available two operational modes of AlignMe: (i) sequence to sequence alignment, taking two sequences in fasta format as input, combining information about each sequence from multiple sources and producing a pair-wise alignment (PW mode); and (ii) alignment of two multiple sequence alignments to create family-averaged hydropathy profile alignments (HP mode). For the PW sequence alignment mode, four different optimized parameter sets are provided, each suited to pairs of sequences with a specific similarity level. These settings utilize different types of inputs: (position-specific) substitution matrices, secondary structure predictions and transmembrane propensities from transmembrane predictions or hydrophobicity scales. In the second (HP) mode, each input multiple sequence alignment is converted into a hydrophobicity profile averaged over the provided set of sequence homologs; the two profiles are then aligned. The HP mode enables qualitative comparison of transmembrane topologies (and therefore potentially of 3D folds) of two membrane proteins, which can be useful if the proteins have low sequence similarity. In summary, the AlignMe web server provides user-friendly access to a set of tools for analysis and comparison of membrane protein sequences. Access is available at http://www.bioinfo.mpg.de/AlignMe

INTRODUCTION

Membrane proteins constitute 20–30% of the proteins in the cell (1–3) and as such play crucial roles in transport and communication across cell membranes, while also being the targets of ∼50% of the medicinal drugs on the market (4–6). Detailed understanding of their molecular mechanisms, and of their interaction with drugs, however, is limited by a dearth of structural data, reflected in the fact that only 2% of the entries in the Protein Data Bank (PDB) are of this class of protein (7–10) and that most of those membrane protein structures are from prokaryotes. Comparative modeling is a valuable alternative under these circumstances, allowing the construction of atomistic models of biomedically relevant mammalian proteins using structurally homologous proteins as templates. Such homology modeling methodologies are strongly dependent on the alignment between the template and target sequences. Accurate sequence alignments are also essential for identifying evolutionary relationships between protein families, even when the structure of a homolog is not available, as well as for detecting whether a known structure is a suitable template for modeling.

The unique environment of a membrane protein compared to a water-soluble protein leads to distinct environmental pressures on their sequences. Such properties can be taken into account in developing sequence alignment software. However, to date, only a few sequence alignment programs have been developed with membrane proteins in mind, or tested using membrane protein datasets, and fewer still have been made widely available via a web server.

We recently developed a software package specifically designed for pair-wise (PW) alignment of membrane proteins that we called AlignMe (11). AlignMe can take into account membrane-specific information in the form of transmembrane predictions, or hydrophobicity scales, while constructing the alignments. In addition, the selection of inputs and the input parameters were optimized for a set of membrane proteins. Finally, AlignMe was tested specifically on alignments of membrane proteins, for which it produced more accurate alignments than a number of other programs, including HHalign (12), MSAProbs (13) and HMAP (14), particularly in cases where the sequence identity of the two proteins is low (11).

*To whom correspondence should be addressed. Tel: +1 301 402 2012; Fax: +1 301 496 6496; Email: lucy.forrest@nih.gov
†The authors wish it to be known that, in their opinion, the first two authors should be considered as Joint First Author.
Present Address:
René Staritzbichler. Translational Oncology (TRON), Johannes Gutenberg University Medical Center Mainz, Mainz 55131, Germany.
Kamil Khafizov. Moscow Institute of Physics and Technology, Moscow 141700, Russian Federation.
Lucy R. Forrest. National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20852, USA.

Published by Oxford University Press on behalf of Nucleic Acids Research 2014.
This work is written by (a) US Government employee(s) and is in the public domain in the US.

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

Other programs specific for membrane proteins include MP-T, which uses environment-specific substitution matrices for generating sequence-to-structure alignments (15). In this context, MP-T is limited to alignments for which one structure is known. As yet, MP-T is not available directly via a web server, except within the Memoir online membrane protein homology modeling workflow (16).

Two other programs specifically designed for membrane proteins, namely PRALINE^{TM} (17) and TM-Coffee (18), are available exclusively as web servers and are mainly intended for constructing multiple-sequence alignments (MSAs); we note that these were not previously compared with AlignMe as they could not be tested with the large data set of PW alignments used for the assessment (11). A number of other MSA web servers not designed specifically for membrane proteins also exist, including Clustal Omega (19), PicXAA (20,21) and PSI-Coffee (18); these also could not be tested on the large data set used (11).

To provide an accessible interface for membrane protein-specific sequence alignments, we developed a web server for AlignMe. The interface allows for two different types of alignment. In the first mode, accurate PW sequence alignments can be generated, such as those required for comparative modeling. The second mode (HP) allows for comparison of family-averaged hydrophobicity profiles by alignment of two sets of sequence homologues; this mode is based on the methodology of Lolkema and Slotboom (22,23), which has been shown to be useful for visualization and comparison of transmembrane topologies (22–26).

The AlignMe software is very flexible, and therefore numerous user-specified options have been made available on the web interface. However, we also provide default, optimized parameters, so that alignments can be computed with minimal human intervention and expertise.

In the following, we describe the available options on the AlignMe web server, including a new fast PW alignment approach. Finally, we briefly mention enhancements that will be incorporated in the future.

THE WEB SERVER

The web server can be accessed from http://www.bioinfo.mpg.de/AlignMe and http://forrestlab.org/AlignMe, and supports all major web browsers (Mozilla Firefox v26, Google Chrome v31, Internet Explorer v11, Safari v7.0). A login to the website is not required but an email address can be provided for users that wish to receive their alignment results via email. In addition, the AlignMe manual and Unix source code are available for download at http://www.bioinfo.mpg.de/AlignMe/download/.

Aside from the home page, and a frequently asked questions page, the web server provides two other tabs that link to either (i) the PW sequence-to-sequence alignment mode or (ii) the hydropathy profile (HP) alignment mode of AlignMe. More details of these two modes are described below.

The current hardware configuration of the web server is a shared resource that includes 208 CPU cores hosted on either Intel Xeon quad-core 3 GHz Woodcrest processors with 8GB shared RAM or 8-core 2.4 GHz Nehalem processors with 36GB shared RAM. A maximum of 208 jobs can be run concurrently.

The AlignMe PW sequence-to-sequence alignment mode

Usage of the PW alignment mode. AlignMe uses the standard Needleman-Wunsch algorithm (in serial C/C++ code) for PW alignments, and the only required input is two protein sequences in standard fasta format. However, AlignMe has been designed to be flexible in handling other input descriptors—essentially anything that reflects the nature of the relationship between two proteins—which can then be used to guide the PW alignment. This flexibility has also been conferred to the web server, so the user can customize all inputs and alignment parameters if required.

Standard parameter sets. For normal usage, we provide four optimized sets of gap penalties and input parameters that lead to accurate alignments under specific circumstances, according to an analysis using BAliBASE reference set 7 of membrane proteins (27,28), compared with a number of other available methods (11); see also Tables 1 and 2. The optimized parameter sets are as follows:

1) AlignMe PST: for aligning distantly related proteins, i.e. with a sequence identity <15%. The inputs consist of a position-specific substitution matrix, PSSM (P), a secondary structure prediction (S) and a transmembrane prediction (T). This strategy resulted in 1.8–7.5% more correctly aligned positions on 206 PW alignments of the BAliBASE ion and pgta families, and significantly smaller shift errors over the whole dataset than the next best method (HMAP or HHalign) (11). To generate the PSSM, a PSI-BLAST search is carried out, which typically takes minutes.
2) AlignMe PS: for aligning low-homology proteins (~15–45% sequence identity). This version is similar to AlignMe PST but omits the membrane prediction. This combination provided the best overall strategy of those tested, giving 6.5% more correctly aligned positions, and smaller shift errors than the next best method (HMAP), over all 15 447 PW BAliBASE alignments (11).
3) AlignMe P: for aligning closely related proteins (>45%). This approach only considers sequence information since it uses only the PSSM and none of the predictions. This strategy results in 4.1% more correctly aligned positions, and significantly smaller shift errors, in alignments of the dtd and photo families of the BAliBASE set 7 than the next best method (HMAP) (11).
4) Fast: for a quick response, the PSI-BLAST search needed for the PST, PS and P versions (11) is best avoided. The web server provides such a fast (<3 s) albeit less accurate mode (see below for details).

Available input descriptors. As for the local version of AlignMe, the web user can define their own alignment parameters and has a choice of input descriptors in three different forms: substitution matrices (BLOSUM62, PHAT, SLIM, VTML or PSI-BLAST PSSM); amino-acid parameters, such as hydrophobicity scales (including values from Eisenberg and Weiss, Hessa *et al.*, Kyte and Doolittle or

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

Table 1. Accuracy of the AlignMe Fast parameter set: Percentage of residues aligned correctly in PW sequence alignments from the BALiBASE reference set 7

	ion	Nat	ptga	7tm	dtd	acr	photo	msl	mean
AlignMe P	38.90	43.50	42.10	42.50	67.10	87.00	**87.90**	**82.50**	61.4
AlignMe PS	45.20	**66.20**	**64.80**	**65.90**	**76.00**	**89.70**	87.60	82.30*	**72.2**
AlignMe PST	**48.10**	58.60	58.78	59.40	71.20	86.30	82.90	76.50	67.7
AlignMe Fast	40.94	53.33	54.86	60.93	66.67	82.13	79.94	77.45	64.5
Number of alignments[a]	1326	1711	1275	8128	1485	903	528	91	
Sequence identity (%)[b]	11.7± 13.8	14.3± 10.8	15.9± 12.1	18.2± 9.7	18.7± 11.5	26.9± 11.3	27.3±1 6.9	35.3± 13.5	

Entries in bold in all tables indicate the highest or best scores in that column. *Values marked with an asterisk in all tables are not significantly different from the highest/best score in a column according to a PW Wilcoxon signed rank test. Mean = mean percentage of correctly aligned residues over averages for eight families. [a]Number of PW alignments. [b]Mean (±standard deviation) of the percentage sequence identity between pairs of alignments in each family. Families are sorted by the average sequence identity.

Table 2. Accuracy of the AligneMe Fast parameter set: Average shift error in PW alignments of the BALiBASE reference set 7

	ion	Nat	ptga	7tm	dtd	acr	photo	msl	mean
AlignMe P	29.92	48.71	33.98	47.58	9.83	1.09*	0.31	0.59	21.50
AlignMe PS	28.83	**2.46**	**3.12**	**3.67**	**1.71**	**0.33**	0.36*	**0.42**	5.11
AlignMe PST	**13.83**	3.24	5.39	11.82	3.46	0.42	**0.31**	0.47	**4.87**
AlignMe Fast	28.18	4.21	10.10	4.27	4.14	0.84	0.58	0.71	6.63
Number of alignments	1326	1711	1275	8128	1485	903	528	91	
Sequence identity (%)	11.7± 13.8	14.3± 10.8	15.9± 12.1	18.2± 9.7	18.7± 11.5	26.9± 11.3	27.3±1 6.9	35.3± 13.5	

The shift error is calculated as the number of positions by which a given residue is misaligned, summed over the length of the alignment and averaged over all alignments. See Table 1 for more details.

Wimley and White) or per-residue profiles, such as a transmembrane prediction from OCTOPUS or a secondary-structure prediction from PSIPRED 3.2. For details see Stamm *et al.* (11). The amino-acid scales can be window-averaged in different ways (e.g. using a triangular form) over any length of window. There are no limits on the number of matrices, scales or profiles that can be combined. However, the different input parameters should be weighted according to the range of values within that scale to prevent bias; details are provided in the user manual.

Outputs. The output from the PW sequence-to-sequence mode includes the PW alignment of the two amino-acid sequences in ClustalW format, the corresponding sequence identity and the percentage of matched positions. Plots are presented that provide a simple representation of the similarity of hydrophobicity, secondary-structure or transmembrane predictions of the two proteins (see Figure 1). In addition, the table of hydrophobicity and/or prediction values for each alignment position is displayed at the bottom of the results page, allowing the user to customize the representation of the data. Finally, a summary of the input parameters used for the alignment is provided. All output files can be downloaded separately or together as a single (archive) file. Results are stored for 14 days on the server and can be retrieved using a Job Identifier (ID), which is provided on the results page.

Accuracy of the Fast PW version of AlignMe. The AlignMe PST, PS and P versions rely on a PSI-BLAST search as well as on secondary structure or transmembrane predictions, and therefore require several minutes to compute. Here, we developed a fast mode of AlignMe to provide the highest possible accuracy for a PW membrane protein alignment without a PSI-BLAST search. Specifically, we selected the combination of inputs that resulted in the best (lowest) alignment difference (AD) score after optimizing against the HOMEP2 membrane protein training data set (Figure 2); the AD score favours matching of residues as well as shorter shift errors (11). The most accurate fast version of AlignMe combines a substitution matrix (VTML (29,30)) with a hydrophobicity scale (from Hessa, White and von Heinje, HWvH (31)); see Figure 2 for details.

We compared the accuracy of the four versions of AlignMe using an independent data set of alignments not used for training, namely the BAliBASE reference set 7 of membrane proteins (Tables 1 and 2). The 'fast' version of AlignMe is somewhat less accurate than the other three ver-

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

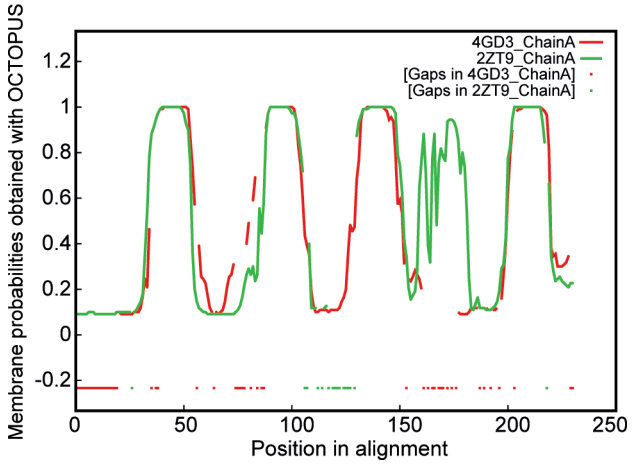

Figure 1. Example output of aligned transmembrane helix probability values for an AlignMe alignment. The sequences of *Escherichia coli* cytochrome b (PDB identifier 4GD3 chain A; red line) and *Nostoc sp. PCC7120* cytochrome b_6f (PDB identifier 2ZT9 chain A; green line) were PW aligned (in PW mode) using the PST parameter set. The transmembrane helix probability from OCTOPUS is plotted for each position in the alignment. Values close to 1 indicate a high likelihood of that sequence being in the membrane. Gaps introduced during alignment are indicated by dots beneath the alignment in the color corresponding to the sequence in which the gap was introduced. In some output plots from AlignMe, a horizontal blue line may be present, indicating that a threshold of probability (usually 0.5) was used to define whether a residue is in a conserved element or not, for selection of the gap penalties. In the AlignMe PST mode, the gap penalty definition threshold is assigned based on the secondary structure prediction not the transmembrane prediction, and is thus not shown in the current figure.

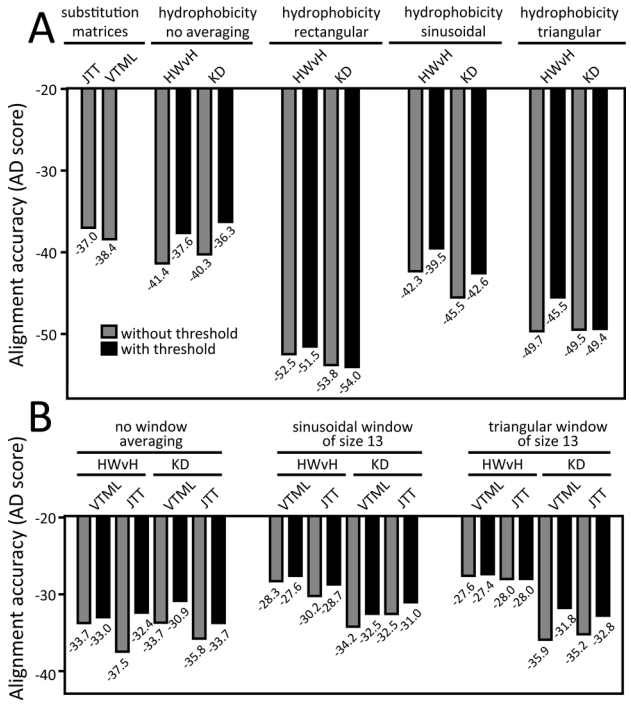

Figure 2. Identification of the optimal combination of inputs for a fast mode of AlignMe, developed for the web server. Alignments generated for the HOMEP2 data set of proteins are compared with the reference alignments using the AD score, according to Stamm *et al.* (11). Alignments were generated using (**A**) a single input, either a substitution matrix or a hydrophobicity scale or (**B**) combinations of these two inputs. Hydropathy scale inputs were aligned without any window-averaging, or using rectangular-, sinusoidal- or triangular-shaped window-averaging of the scores along the sequence. In some cases (black bars), a threshold of hydrophobicity = 0 was used to assign different gap penalties in either conserved (hydrophobic) or variable (polar) regions of the sequence. Alignment accuracy improved (less negative AD scores) using combinations of matrices with hydrophobicity scales. The alignments most similar to the references were obtained using the VTML matrix and the HWvH hydrophobicity scale averaged over a 13-residue long triangular window, while applying different gap penalties in different regions of the alignment according to the aforementioned threshold. The optimized gap penalty values for these settings were: $p_o^{\text{above}} = 11.0$, $p_e^{\text{above}} = 4.4$, $p_o^{\text{below}} = 18.0$, $p_e^{\text{below}} = 1.3$, $p_o^{\text{terminal}} = 13.1$ and $p_e^{\text{terminal}} = 0.9$.

sions on BAliBASE, but appears to provide a compromise between the PST, PS and P parameter sets, depending on the sequence identity range. Due to its speed, this fast mode therefore provides a useful first-pass approach, e.g. to approximate the sequence identity of a pair of sequences before selecting the AlignMe PST, PS or P mode to obtain the most accurate alignment.

Alignment of Family-averaged HPs using two MSAs (HP mode)

Usage of the HP mode. The transmembrane topology of a membrane protein is reflected in the shape of the hydrophobicity of its sequence, with strong peaks in the most hydrophobic transmembrane segments. Since hydrophobicity is typically conserved in transmembrane helices during evolution, these profiles can contain similar global features even in very distantly related proteins. Although lacking in detail, as well as a meaningful significance score, comparison of HPs can provide an intuitive overview of the similarities between transmembrane topologies of two proteins (22–24). Averaging each of the input profiles over a set of sequence homologues, in a so-called family-averaged HP, can smooth out noise and sequence-specific detail, making comparisons clearer (22–24). To date, the ability to generate these aligned HPs has not been readily available to the community. On the AlignMe web server, we provide a simple interface to such comparisons via the 'alignment of two MSAs' tab.

Inputs. To obtain family-averaged HPs, the sequence homologues must be supplied in the form of MSAs, one for each input family, although individual HPs can also be analyzed by providing a single sequence. The sequences within a provided MSA must be aligned (i.e. including gaps so that all entries are the same length) and in fasta format. Thus, the user would typically first carry out a PSI-BLAST search for sequence homologues of each query protein, and align each set of homologues with a MSA program, such as Clustal Omega (19). There is no limit on the size (length or number of sequences) of the membrane protein MSAs that can be provided when using the HP mode. AlignMe then calculates the average hydrophobicity value for every position (i.e. column) in the MSA to produce a family-averaged HP.

Provided parameter sets. The pre-defined parameter set provided on the web server for HP alignments was obtained previously by a systematic optimization procedure designed

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

to identify five-transmembrane domains in a hydropathy-profile search, and adjusted slightly for alignment rather than detection (24). Specifically, values from the HWvH hydrophobicity scale (31) are smoothed using a 15-residue long, triangular sliding window, using the following gap penalties: $p_o^{\text{above}} = 2.5$, $p_e^{\text{above}} = 1.0$, $p_o^{\text{below}} = 1.0$, $p_e^{\text{below}} = 0.85$, $p_o^{\text{terminal}} = 0.25$ and $p_e^{\text{terminal}} = 0.25$, where the hydrophobicity threshold used to assign 'above' and 'below' was –0.5. However, users have a choice of hydrophobicity scale and/or gap penalties.

An additional parameter in the profile-to-profile alignments is the fraction of allowed gaps, which defines the number of low-confidence columns in the input MSA that will be considered. The default value is 0.5, meaning that columns containing >50% gaps are ignored.

Outputs. The two averaged profiles from the input MSAs are aligned to one another by treating them as standard profiles, as in the PW mode. Alignments of two MSAs are presented in a hydropathy plot, and in a PW sequence alignment of the first sequence of each of the two MSAs in ClustalW format, as well as in a table of the hydropathy values. Gaps in the alignment that were present in the original MSAs are represented by a '.' symbol, whereas gaps introduced during the alignment of the averaged HPs are indicated with a '-' symbol. Input parameters and results can be downloaded separately or all in a single file.

Example applications of HP mode alignments. HPs have been used in a number of studies to assess the topological similarity of two proteins with very low or undetectable sequence similarities. For example, evolutionary relationships have been illustrated between neurotransmitter:sodium symporters, sodium:solute symporters and members of the amino-acid/polyamine/organocation superfamily (32); between internal structural repeats of transporters (24); between the 2-hydroxycarboxylate transporters (2HCT) and so-called ESS transporter families (33,34); between a multidrug and toxin extrusion transporter and the inner membrane flippase Wzx (35); as well as between the SLC13 and SLC34 transporter families (26). The same basic approach was also used in the MemGen classification of numerous secondary transporters (36–38), in a comparison of members of the sodium-phosphate transporter family NaPi-II (25), and to identify a putative ancestral half-transporter (24).

CONCLUSION

The AlignMe web server provides a user-friendly interface for a set of sequence alignment approaches specifically tuned to membrane proteins. In the PW option the user can readily compute PW alignments suitable, e.g. for homology modeling. To estimate the sequence identity of the protein pairs, the user might first use the fast mode, before running the slower, but more accurate mode most suited to the detected sequence identity range.

The second functionality provided, namely of HP alignments, provides for the first time a user-friendly interface for comparing the topological nature of two membrane proteins, as originally developed by Lolkema and Slotboom (22,23).

Future developments will include the ability to accurately align β-barrel proteins, which exhibit distinct properties, and fold classification, to score a sequence against all known membrane protein structures.

ACKNOWLEDGEMENT

We thank members of the Forrest lab for beta-testing the web server, and the 'infrastructure for bioinformatic applications' team from the Computing Centre (RZG) of the Max Planck Society, Garching, Germany for servers and assistance.

FUNDING

DFG (German Research Foundation), Collaborative Research Center 807 'Transport and Communication across Biological Membranes'; Max Planck Society; Intramural Research Program of the National Institutes of Health, NINDS. Source of Open Access funding: National Institutes of Health, National Institute of Neurological Disorders and Stroke (NIH-NINDS).

Conflict of interest statement. None declared.

REFERENCES

1. Jones,D.T. (1998) Do transmembrane protein superfolds exist? *FEBS Lett.*, **423**, 281–285.
2. Krogh,A., Larsson,B., von Heijne,G. and Sonnhammer,E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
3. Nugent,T. and Jones,D.T. (2009) Transmembrane protein topology prediction using support vector machines. *BMC Bioinform.*, **10**, 159.
4. Russ,A.P. and Lampel,S. (2005) The druggable genome: an update. *Drug Discov. Today*, **10**, 1607–1610.
5. Drews,J. (2000) Drug discovery: a historical perspective. *Science*, **287**, 1960–1964.
6. Hopkins,A.L. and Groom,C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.*, **1**, 727–730.
7. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig, H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235-242.
8. Tusnady,G.E., Dosztanyi,Z. and Simon,I. (2004) Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics*, **20**, 2964–2972.
9. Tusnady,G.E., Dosztanyi,Z. and Simon,I. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res.*, **33**, D275–D278.
10. Kozma,D., Simon,I. and Tusnady,G.E. (2013) PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.*, **41**, D524–D529.
11. Stamm,M., Staritzbichler,R., Khafizov,K. and Forrest,L.R. (2013) Alignment of helical membrane protein sequences using AlignMe. *PloS One*, **8**, e57731.
12. Söding,J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
13. Liu,Y., Schmidt,B. and Maskell,D.L. (2010) MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. *Bioinformatics*, **26**, 1958–1964.
14. Tang,C.L., Xie,L., Koh,I.Y., Posy,S., Alexov,E. and Honig,B. (2003) On the role of structural information in remote homology detection and sequence alignment: new methods using hybrid sequence profiles. *J. Mol. Biol.*, **334**, 1043–1062.
15. Hill,J.R. and Deane,C.M. (2013) MP-T: improving membrane protein alignment for structure prediction. *Bioinformatics*, **29**, 54–61.

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

16. Ebejer,J.P., Hill,J.R., Kelm,S., Shi,J. and Deane,C.M. (2013) Memoir: template-based structure prediction for membrane proteins. *Nucleic Acids Res.*, **41**, W379–W383.
17. Pirovano,W., Feenstra,K.A. and Heringa,J. (2008) PRALINETM: a strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics*, **24**, 492–497.
18. Chang,J.M., Di Tommaso,P., Taly,J.F. and Notredame,C. (2012) Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinform.*, **13(Suppl.4)**, S1.
19. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
20. Sahraeian,S.M. and Yoon,B.J. (2011) PicXAA-Web: a web-based platform for non-progressive maximum expected accuracy alignment of multiple biological sequences. *Nucleic Acids Res.*, **39**, W8–W12.
21. Sahraeian,S.M. and Yoon,B.J. (2010) PicXAA: greedy probabilistic construction of maximum expected accuracy alignment of multiple sequences. *Nucleic Acids Res.*, **38**, 4917–4928.
22. Lolkema,J.S. and Slotboom,D.J. (1998) Estimation of structural similarity of membrane proteins by hydropathy profile alignment. *Mol. Membr. Biol.*, **15**, 33–42.
23. Lolkema,J.S. and Slotboom,D.J. (1998) Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins. *FEMS Microbiol. Reviews*, **22**, 305–322.
24. Khafizov,K., Staritzbichler,R., Stamm,M. and Forrest,L.R. (2010) A study of the evolution of inverted-topology repeats from LeuT-fold transporters using AlignMe. *Biochemistry*, **49**, 10702–10713.
25. Forster,I.C., Kohler,K., Biber,J. and Murer,H. (2002) Forging the link between structure and function of electrogenic cotransporters: the renal type IIa Na^+/Pi cotransporter as a case study. *Prog. Biophys. Mol. Biol.*, **80**, 69–108.
26. Fenollar-Ferrer,C., Patti,M., Knöpfel,T., Werner,A., Forster,I.C. and Forrest,L.R. (2014) Structural fold and binding sites of the human Na^+-phosphate cotransporter NaPi-II. *Biophys. J.*, **106**, 1268–1279.
27. Bahr,A., Thompson,J.D., Thierry,J.C. and Poch,O. (2001) BAliBASE (Benchmark Alignment dataBASE): enhancements for repeats, transmembrane sequences and circular permutations. *Nucleic Acids Res.*, **29**, 323–326.
28. Thompson,J.D., Plewniak,F. and Poch,O. (1999) BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs. *Bioinformatics*, **15**, 87–88.
29. Müller,T., Spang,R. and Vingron,M. (2002) Estimating amino acid substitution models: a comparison of Dayhoff's estimator, the resolvent approach and a maximum likelihood method. *Mol. Biol. Evol.*, **19**, 8–13.
30. Müller,T. and Vingron,M. (2000) Modeling amino acid replacement. *J. Comput. Biol.*, **7**, 761–776.
31. Hessa,T., Kim,H., Bihlmaier,K., Lundin,C., Boekel,J., Andersson,H., Nilsson,I., White,S.H. and von Heijne,G. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, **433**, 377–381.
32. Lolkema,J.S. and Slotboom,D.J. (2008) The major amino acid transporter superfamily has a similar core structure as Na^+-galactose and Na^+-leucine transporters. *Mol. Membr. Biol.*, **25**, 567–570.
33. Dobrowolski,A., Sobczak-Elbourne,I. and Lolkema,J.S. (2007) Membrane topology prediction by hydropathy profile alignment: membrane topology of the Na(+)-glutamate transporter GltS. *Biochemistry*, **46**, 2326–2332.
34. ter Horst,R. and Lolkema,J.S. (2012) Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity. *Biochim. Biophys. Acta*, **1818**, 72–81.
35. Islam,S.T., Fieldhouse,R.J., Anderson,E.M., Taylor,V.L., Keates,R.A., Ford,R.C. and Lam,J.S. (2012) A cationic lumen in the Wzx flippase mediates anionic O-antigen subunit translocation in Pseudomonas aeruginosa PAO1. *Mol. Microbiol.*, **84**, 1165–1176.
36. Lolkema,J.S. and Slotboom,D.J. (2003) Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis. *J. Mol. Biol.*, **327**, 901–909.
37. Lolkema,J.S. and Slotboom,D.J. (2005) Sequence and hydropathy profile analysis of two classes of secondary transporters. *Mol. Membr. Biol.*, **22**, 177–189.
38. Sobczak,I. and Lolkema,J.S. (2005) Structural and mechanistic diversity of secondary transporters. *Curr. Opin. Microbiol.*, **8**, 161–167.

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015